The goal of this problem is to show that there are two equivalent expressions for the residual sum of squares in linear regression.
Let our training data be $\{(\mathbf{x_1}, y_1), \ldots, (\mathbf{x_N}, y_N)\}$, where the vector $\mathbf{x_n} \in \mathbb{R}^D$ includes the D features and $y_n \in \mathbb{R}$ is the label of sample n.

The training data can be also written in a matrix/vector notation as:

$$\mathbf{X} = \begin{bmatrix} 1 & \mathbf{x_1}^T \\ & \vdots \\ 1 & \mathbf{x_N}^T \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1D} \\ & & \vdots & & \\ 1 & x_{N1} & x_{12} & \ldots & x_{ND} \end{bmatrix} \in \mathbb{R}^{N \times (D+1)} \text{ and } \mathbf{y} = [y_1, \ldots, y_N]^T \in \mathbb{R}^N$$

where x_{nd} is the d^{th} feature of sample n.

Let's also assume that the weight of the linear regression model is written as $\mathbf{w} = [w_0,\ w_1,\ \ldots,\ w_D]$, where w_0 is the bias.

Show that the following expressions of the RSS error are equivalent:

$$RSS(\mathbf{w}) = \sum_{n=1}^{N} \left[y_n - \left(w_0 + \sum_{d=1}^{D} w_d x_{nd} \right) \right]^2$$

$$RSS(\mathbf{w}) = (\mathbf{y} - \mathbf{X}\mathbf{w})^T (\mathbf{y} - \mathbf{X}\mathbf{w})$$

Let $\mathbf{x} = [x_1, x_2] \in \mathbb{R}^2$ be a 2-dimensional vector.

(1) Compute the first derivative and Hessian of the function f.

(2) Show that the function $f(\mathbf{x}) = (x_1 - 3x_2)^2$ is convex by showing that its Hessian is positive semi-definite.

(3) Show that the function $f(\mathbf{x}) = (x_1 - 3x_2)^2$ is convex by showing that the eigenvalues of its Hessian are non-negative.